DOCUMENT RESUME

ED 315 444 TM 014 470

TITLE Maryland Functional Mathematics Test-II: Fall 1987 Administration. Report M-87. Technical Report.
INSTITUTION Maryland State Dept. of Education, Baltimore. Div. of Instruction.
PUB DATE Aug 88
NOTE 19p.; Some tables contain small type.
PUB TYPE Statistical Data (110) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Grade 9; *Graduation Requirements; High Schools; *High School Students; Mathematics Achievement; Mathematics Skills; *Mathematics Tests; Public Schools; *State Programs; Testing Programs; *Test Results
IDENTIFIERS *Maryland Functional Mathematics Test

ABSTRACT

The Maryland Functional Mathematics Test, Level II (MFMT-II) assesses Maryland students' mastery of functional mathematics skills that are described in the Declared Competencies Index published in 1980 by the Maryland State Department of Education. The test is also used to certify students for graduation from high school and to identify students in need of assistance with their math skills. Forms 260 through 269 of the 1987 edition of the MFMT-II were administered to students in grades 9 through 12 during the period October 19 to November 6, 1987. Students in all 24 school districts took the MFMT-II for the first time in grade 9; those in grades 10 through 12 who took it had failed in previous years or were transfers. In all, 46,635 students took the test in 1987. The graduating class of 1987 was the first required to pass the MFMT-II to receive a high school diploma. Separate sections of this paper describe and report the development procedures, technical characteristics, and student performance results of the 1987 administration of the MFMT-II. More detailed information about testing methods, reliability, and validity is to be included in the technical manual for the testing program. In 1987, the percentage of students who passed on the first attempt was higher than in 1986, with a 1.3 point increase in the mean scale score (from 351.5 in 1986 to 352.8 in 1987). Nine tables contain the test results and passing percentages by grade. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED315444

TECHNICAL REPORT

MARYLAND FUNCTIONAL MATHEMATICS TEST-II
FALL 1987 ADMINISTRATION
REPORT M-87
AUGUST 1988

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. A. Clenson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Maryland State Department of Education
Division of Instruction
Program Assessment, Evaluation, and Instructional Support Branch
200 W. Baltimore Street
Baltimore, Maryland 21201
(301) 333-2369

MFTP

MARYLAND FUNCTIONAL TESTING PROGRAM

A COMPONENT OF PROJECT BASIC FOR:

* **ASSESSMENT OF BASIC AND CITIZENSHIP SKILLS**
* **IDENTIFICATION FOR APPROPRIATE ASSISTANCE**
* **CERTIFICATION FOR GRADUATION**

BEST COPY AVAILABLE

TM014470

Technical Report

MARYLAND FUNCTIONAL MATH TEST-II
Fall 1987 Administration

Maryland State Department of Education
Division of Instruction
Program Assessment, Evaluation, and Instructional Support Branch

August 1988

The purpose of the Maryland Functional Math Test-II is threefold. First, the test has been designed to assess Maryland students' mastery of functional math skills that are described in the Declared Competencies Index (DCI) published by the Maryland State Department of Education (MSDE) in 1980. In addition, the test is used to certify students for high school graduation. Finally, the test is used to identify students who should participate in appropriate assistance in order to improve their math skills.

Forms 260-269 of the 1987 edition of the Maryland Functional Math Test, Level II (MFMT-II) were administered in Maryland public schools to students in grades 9-12 during the period October 19-November 6, 1987. Students in all of the state's 24 school districts took the MFMT-II for their first time in ninth grade. Grade 10, 11, and 12 students who took the 1987 MFMT-II include members of the graduating classes of 1988 and beyond who failed the test in previous years as well as transfer students. The graduating class of 1987 was the first class required to pass the MFMT-II in order to receive a Maryland High School diploma.

The remainder of this report contains separate sections describing and reporting the development procedures, technical characteristics, and student performance results of the 1987 edition of the MFMT-II. For more detailed descriptions of the procedures used in test development, as well as more technical descriptions of psychometric methods used to evaluate functional tests, refer to the Maryland Functional Testing Program (MFTP) Technical Manual currently in production.

DEVELOPMENT

Declared Competencies Index and the MFMT-II

All schools are required to teach the content areas described in the Declared Competencies Index (DCI). The DCI provides the basis for the test specifications and objectives assessed by the MFMT-II. In December 1981, staff from the Division of Instruction (DOI) and Project Basic produced the Project Basic Instructional Guide, Volume V: Functional Mathematics (MSDE, 1981) designed to help teachers and

supervisors develop instructional activities to prepare students for the MFMT-II.

Test Specifications

The Test Specifications (MSDE, 1982) for the MFMT-II define the eligible content for the seven domains of the MFMT-II that include: number concepts, whole number operations, mixed number/fraction operations, decimal operations, measurement, using data, and problem solving. This document describes test specifications that guide item writers in generating new items for both field testing and operational use on subsequent forms of the MFMT-II. It is periodically reviewed and revised according to the general procedures described more fully in the MFTP Technical Manual. The test specifications committee for the MFMT-II did not meet prior to the development of the 1987 test, and no revisions were proposed or made to the specifications during that time.

Item Development

When necessary, the item writing committee for the MFMT-II convenes to generate new items to assess the objectives in each of the seven domains in the test. Items are checked for readability, item/objective match, and conformance with test specifications. They are reviewed, revised, and field tested prior to inclusion in the MFMT-II item bank. Although no new items were generated by the committee for the 1987 MFMT-II, several existing items from the item bank were revised, reviewed and re-field tested on the 1987 MFMT-II.

Item Review

In order to be included on an edition of a test in the Maryland Functional Testing Program, items must pass content, bias, psychometric, and editorial reviews that are described in detail in the MFTP Technical Manual. The operational and field test items included in the 1987 MFMT-II were submitted to these reviews.

Test Forms

Each year, ten forms of the MFMT-II are administered. All ten forms have the same 70 operational items that count toward a student's total score on the test. Each form of the test also includes seven field test items. This format allows the Program Assessment, Evaluation, and Instructional Support Branch (PAEISB) to gather field test data on a total of 70 items each year. For further information on field test procedures and the purposes of field testing, consult the MFTP Technical Manual.

TECHNICAL CHARACTERISTICS

Reliability

The reliability of the MFMT-II is evaluated using two statistical indicators: internal consistency and classification consistency coefficients. Scores on criterion-referenced tests (CRTs) such as the MFMT-II are used to infer the degree to which students have mastered a domain of knowledge and skill. To make valid inferences from the MFMT-II to the domain of mathematics knowledge and skill defined by the DCI, items must measure only this knowledge and skill. Internal consistency reliability coefficients provide a measure of the extent to which the items on a test are homogeneous, thereby providing one way of describing test reliability. One internal consistency reliability coefficient is the Kuder-Richardson formula number 20, or KR20 (see, for example, Crocker & Algina, 1986; Gronlund, 1982.)

The MFMT-II is accompanied by a passing score adopted by the Maryland State Board of Education. This score resulted from standard setting procedures which included recommendations from groups of citizens, teachers, administrators, parents, and students from around the state, as well as student field test performance data from the fall of 1984. This passing score corresponds to answering approximately 80% of the operational items correctly on a given test and signifies the level of performance required for graduation certification. The standard setting procedures for the MFMT-II are discussed in the MSDE publication entitled Setting a Standard for Standard Setting, (Chafin, 1983). Students whose MFMT-II score is above the passing score are considered to have mastered this set of functional skills, and students scoring below the passing score are considered not to have mastered the content and, therefore, in need of remedial instruction. The reliability or consistency with which the MFMT-II classifies students into these two categories is essential for drawing valid, believable conclusions about a student's mastery of functional mathematics. An indicator of the consistency of classification is an adaptation of a statistic developed by Subkoviak (in Berk, 1984, pp.278-280), and referred to here as the PG coefficient. The reliability and classification consistency of the 1987 MFMT-II are discussed in the following sections.

Table 1

Internal Consistency (KR20) and Classification Consistency (PG) Coefficients

Individual test forms (77 items each form*)

Form	No. Students	KR20
260	4,538	.86
261	4,480	.86
262	4,551	.86
263	4,536	.87
264	4,689	.87
265	4,502	.86
266	4,385	.87
267	4,475	.86
268	4,420	.86
269	4,581	.86

All forms, all students (N=46,635)

Domain	No. Items	KR20	PG
1. Number Concepts	10	.74	.84
2. Whole Number Operations	10	.60	.62
3. Mixed Number/Fraction Operations	10	.85	.85
4. Decimal Operations	10	.72	.85
5. Measurement	10	.69	.82
6. Using Data	10	.73	.79
7. Problem Solving	10	.75	.81
Total Test	70	.93	.90

*KR20 coefficients are not available at this time for the 70 operational items

Internal Consistency. Table 1 contains the internal consistency of the 1987 MFMT-II forms and items across forms. KR20 coefficients for the individual forms have remained stable and at a high level over the last three years. KR20 coefficients for the 10 forms of the MFMT-II ranged from .86 to .87. Although each domain of the test has only ten items, KR20 coefficients are satisfactory for all domains, ranging from .61 to .85 for the 1987 test, and have remained fairly stable over the last three years. The KR20 coefficient for the total test is higher than those for the individual domains because domain KR20 coefficients are calculated on fewer items than for the total test. KR20s are lower bound estimates of test reliability.

Classification consistency. Table 1 indicates that the classification consistency of the MFMT-II is excellent: .90 for the students who took the test in 1987. This coefficient reveals that approximately 90% of the students who took the test would be classified in the same way (as a passer or failer) on a second concurrent administration of the test. The PG of .90 indicates that about 10 percent of the students who took the test might have been misclassified as passers or failers. It is likely that students who may have been classified incorrectly are those whose total scale score was just above or just below the passing score of 340.

Validity

The primary aim in developing, administering, scoring, and reporting the MFMT-II is to provide accurate decisions about students' mastery of mathematics skills. The first condition necessary to assure the validity of any test is the demonstration of acceptable levels of reliability and classification consistency as described in the previous section. Evidence of the validity of the MFMT-II includes information about content validity, construct validity, and freedom from bias. The MFTP Technical Manual will include descriptions of the procedures used to help assure that the MFMT-II is administered according to prescribed procedures, methods used to determine content and construct validity, and procedures used to assist in determining that items are free of bias.

Equivalence of Forms

The issue of equity, or fairness, is of major importance to any test that has a designated passing score and is used to make decisions concerning high school graduation. Students who take the MFMT-II must be assured that the test they are taking is no more easy or difficult to pass than other editions or forms of the test. Thus, equivalence of test forms refers to equivalence of the difficulty of passing a particular form of the test.

Items are selected for inclusion on a given edition of the MFMT-II to help ensure the equivalence of current forms to

previous editions. Statistical pre-equating analysis is performed before the test is administered to ensure that the forms are equivalent in difficulty to one another and to previous editions. This analysis also provides a way to ensure comparability of the passing standard. The details of PAEISB's Rasch equating procedure can be found in the MFTP Technical Manual.

Table 2 shows raw scores (number of items correct) and their corresponding equated scale, domain, and percent correct scores and standard errors. Students who correctly answered 57 items on the 1987 MFMT-II achieved a scale score of 340, which is the lowest possible passing scale score in 1987. This passing raw score is one point higher than the lowest possible passing raw score of 56 in 1986. This adjustment in passing score was necessary to assure that the 1987 MFMT-II was the same level of difficulty as the 1986 MFMT-II.

Table 3 contains raw scores, equated scale scores, domain scores, and percent correct scores and standard errors for each domain. To illustrate the use of this table, the panel for domain 1 indicates that a student who correctly answered 8 of 10 items, equivalent to a scale score of 342, can be considered to have mastered the domain. However, pass-fail status is not determined by performance within an individual domain: a student may perform poorly on a given domain, yet pass the test by performing well in other domains. One reason for determining pass-fail status only at the total test level is that longer tests are generally more reliable than shorter tests. This relationship between test length and test reliability is reflected in the scale score standard errors (SEss) for domain one. A scale score of 342, based on ten items in the domain, has a standard error of 16--considerably higher than the standard error of 7 for a scale score of 342 based on 70 items (See Table 2).

Item Bias Analysis

The procedures used to detect item bias during the development phase of the MFMT-II will be described in detail in the MFTP Technical Manual. In addition to these judgmental reviews, items are subjected to statistical analysis to determine any sex and race bias after the test has been administered. An adaptation of the delta-plot method (Angoff, 1982, pp. 96-106) is used to identify items that may be biased so that they can be resubmitted for review prior to inclusion in a student's score on the MFMT-II.

Tables 4 and 5 contain the 1987 MFMT-II delta differences for males versus females and blacks versus whites, respectively. A delta difference of positive or negative 3.5 or higher indicates an item that might be biased and should be reviewed before final computation of student scores. Non-significant differences between deltas for each group are likely due to random fluctuations in sub-group performances and measurement

Table 2

Lookup Table For The Total Test

DOMAIN PROCEDURE

NUMBER OF ITEMS = 70 MEAN DIFFICULTY = .299 STD. DEVIATION = 1.167

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	0	0	-5.577	1.472	185	29	1.13
1	2	1	-4.613	.982	205	20	1.83
2	3	2	-3.877	.735	219	15	2.55
3	5	4	-3.426	.615	228	12	3.01
4	7	5	-3.094	.535	235	11	3.32
5	9	7	-2.827	.490	240	10	3.61
6	11	8	-2.602	.458	245	9	3.84
7	13	10	-2.406	.429	249	9	4.03
8	14	11	-2.231	.407	252	8	4.19
9	16	12	-2.073	.389	256	8	4.32
10	17	14	-1.926	.370	258	7	4.39
11	19	15	-1.791	.359	261	7	4.50
12	21	17	-1.665	.349	264	7	4.60
13	23	18	-1.545	.340	266	7	4.68
14	23	20	-1.432	.332	268	7	4.75
15	26	21	-1.324	.324	271	6	4.81
16	27	22	-1.221	.318	273	6	4.86
17	29	24	-1.121	.312	275	6	4.90
18	30	25	-1.025	.307	276	6	4.93
19	31	27	-.933	.302	278	6	4.96
20	33	28	-.843	.298	280	6	4.99
21	34	30	-.755	.294	282	6	5.01
22	37	31	-.669	.291	284	6	5.02
23	38	32	-.586	.288	285	6	5.01
24	39	34	-.504	.284	287	6	5.02
25	41	35	-.424	.281	289	6	5.03
26	42	37	-.345	.279	290	6	5.04
27	44	38	-.267	.277	292	6	5.04
28	45	40	-.191	.276	293	6	5.04
29	46	41	-.115	.274	295	5	5.03
30	48	42	-.040	.273	296	5	5.02
31	49	44	.034	.272	298	5	5.01
32	50	45	.108	.271	299	5	5.00
33	52	47	.181	.270	301	5	4.98
34	53	48	.254	.270	302	5	4.96
35	55	50	.327	.270	304	5	4.95
36	56	51	.400	.269	305	5	4.92
37	56	52	.472	.269	306	5	4.90
38	59	54	.545	.270	308	5	4.87
39	59	55	.618	.270	309	5	4.85
40	61	57	.691	.271	311	5	4.81
41	62	58	.765	.272	312	5	4.78
42	63	60	.839	.273	314	5	4.74
43	64	61	.914	.274	315	5	4.71
44	67	62	.989	.276	317	6	4.66
45	67	64	1.066	.277	318	6	4.62
46	69	65	1.143	.279	320	6	4.57
47	70	67	1.222	.281	321	6	4.52
48	71	68	1.302	.284	323	6	4.48
49	73	70	1.384	.286	325	6	4.40
50	74	71	1.468	.291	326	6	4.36
51	75	72	1.554	.295	328	6	4.30
52	76	74	1.643	.300	330	6	4.23
53	78	75	1.734	.304	332	6	4.16
54	79	77	1.828	.310	334	6	4.08
55	80	78	1.926	.316	336	6	4.00
56	81	80	2.029	.323	338	6	3.91
57	82	81	2.136	.331	340	7	3.81
58	84	82	2.249	.340	342	7	3.70
59	85	84	2.369	.350	344	7	3.59
60	87	85	2.497	.362	347	7	3.46
61	88	87	2.635	.375	350	7	3.32
62	89	88	2.787	.387	353	8	3.22
63	90	90	2.953	.419	356	8	3.06
64	92	91	3.141	.447	360	9	2.88
65	93	92	3.357	.481	364	10	2.67
66	94	94	3.614	.528	369	11	2.42
67	95	95	3.936	.607	376	12	2.17
68	97	97	4.377	.720	385	15	1.81
69	98	98	5.104	.980	399	20	1.29
70	99	100	6.084	1.470	418	29	.82

Table 3

Equated Raw, Domain and Scale Scores and Standard Errors For Each Domain

DOMAIN 1 Number Concepts

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = -1.255 STD. DEVIATION = 1.095

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	3	0	-4.978	1.625	197	33	6.23
1	11	10	-3.802	1.083	221	22	10.57
2	23	20	-2.925	.831	239	17	13.51
3	34	30	-2.316	.738	251	15	14.95
4	44	40	-1.798	.705	261	14	15.56
5	56	50	-1.308	.701	271	14	15.33
6	66	60	-.804	.722	281	14	14.29
7	76	70	-.247	.775	292	16	12.45
8	85	80	.427	.875	308	18	9.73
9	93	90	1.391	1.126	325	23	6.29
10	97	100	2.658	1.688	350	34	3.18

DOMAIN 2 Whole Number Operations

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = .704 STD. DEVIATION = .774

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	5	0	-2.885	1.623	239	32	7.95
1	14	10	-1.713	1.082	263	22	10.89
2	25	20	-.840	.827	280	17	12.31
3	34	30	-.240	.732	292	15	12.88
4	43	40	.259	.687	302	14	12.98
5	52	50	.719	.672	311	13	12.83
6	61	60	1.174	.681	320	14	12.47
7	70	70	1.662	.717	330	14	11.85
8	79	80	2.244	.815	342	16	11.03
9	88	90	3.095	1.071	359	21	9.39
10	95	100	4.242	1.607	382	32	8.24

DOMAIN 3 Mixed Number/Fraction Operations

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = .408 STD. DEVIATION = .345

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	3	0	-2.954	1.587	238	32	5.98
1	10	10	-1.835	1.058	260	21	9.77
2	20	20	-1.013	.796	277	16	12.17
3	30	30	-.462	.692	288	14	13.37
4	39	40	-.009	.654	297	13	14.19
5	48	50	.407	.641	305	13	14.50
6	57	60	.823	.653	313	13	14.42
7	67	70	1.274	.691	322	14	13.80
8	77	80	1.824	.795	333	16	12.74
9	88	90	2.644	1.057	350	21	10.31
10	95	100	3.761	1.585	372	32	8.28

DOMAIN 4 Decimal Operations

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = .631 STD. DEVIATION = 1.172

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	5	0	-3.258	1.681	232	33	8.63
1	15	10	-2.029	1.107	256	22	13.21
2	29	20	-1.100	.862	275	17	14.89
3	41	30	-.441	.772	288	15	14.95
4	52	40	.119	.731	299	15	14.26
5	63	50	.642	.718	310	14	13.18
6	71	60	1.103	.728	320	15	11.77
7	79	70	1.716	.781	331	15	10.07
8	86	80	2.381	.852	344	17	8.17
9	93	90	3.288	1.097	362	22	5.75
10	97	100	4.472	1.645	386	33	3.15

Table 3 (cont.)

Equated Raw, Domain and Scale Scores and Standard Errors For Each Domain

DOMAIN 5 Measurement

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = 1.151 STD. DEVIATION = 1.363

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	6	0	-3.115	1.789	235	35	9.16
1	17	10	-1.724	1.179	263	24	13.00
2	31	20	-.654	.923	284	18	13.86
3	43	30	.092	.812	299	16	13.39
4	53	40	.697	.750	311	15	12.51
5	62	50	1.238	.724	322	14	11.59
6	70	60	1.759	.726	332	15	10.64
7	77	70	2.307	.759	343	15	9.59
8	85	80	2.947	.851	356	17	8.34
9	91	90	3.859	1.102	374	22	6.43
10	96	100	5.073	1.653	398	33	3.93

DOMAIN 6 Using Data

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = -.081 STD. DEVIATION = .885

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	4	0	-3.695	1.615	223	32	7.03
1	12	10	-2.536	1.077	246	22	9.98
2	23	20	-1.672	.824	264	16	11.23
3	31	30	-1.074	.729	276	15	11.83
4	40	40	-.568	.696	286	14	12.33
5	49	50	-.092	.687	295	14	12.66
6	57	60	.386	.700	305	14	12.86
7	66	70	.802	.742	315	15	12.84
8	76	80	1.514	.833	327	17	12.29
9	87	90	2.393	1.084	345	22	10.54
10	95	100	3.567	1.625	368	33	8.75

DOMAIN 7 Problem Solving

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = .540 STD. DEVIATION = .892

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	4	0	-2.955	1.602	238	32	7.15
1	12	10	-1.815	1.068	261	21	10.92
2	24	20	-.971	.811	278	16	12.68
3	33	30	-.396	.712	289	14	13.33
4	43	40	.084	.675	299	14	13.67
5	52	50	.531	.665	308	13	13.64
6	62	60	.980	.679	317	14	13.31
7	69	70	1.468	.723	326	14	12.65
8	79	80	2.054	.818	338	16	11.44
9	89	90	2.813	1.075	355	21	9.23
10	96	100	4.068	1.612	378	32	5.72

Table 4

Delta Values for Bias Analysis: Males vs. Females

OPERATIONAL SEQUENCE NUMBER	P-VALUE G1	P-VALUE G2	Z-VALUE G1	Z-VALUE G2	DELTA G1	DELTA G2	DELTA (G1-G2)
I001	.98	.98	1.75	2.05	20.00	21.20	-1.20
I002	.95	.97	1.65	1.88	19.60	20.52	-.92
I003	.98	.99	2.05	2.33	21.20	22.32	-1.12
I004	.95	.97	1.65	1.88	19.60	20.52	-.92
I005	.94	.98	1.55	1.75	19.20	20.00	-.80
I006	.98	.98	2.05	2.05	21.20	21.20	.00
I007	.87	.87	1.13	1.13	17.52	17.52	.00
I008	.92	.95	1.41	1.65	18.64	19.60	-.96
I009	.79	.84	.81	.99	16.24	16.96	-.72
I010	.81	.85	.88	1.04	16.52	17.16	-.64
I011	.92	.95	1.41	1.65	18.64	19.60	-.96
I012	.84	.88	.99	1.17	16.96	17.68	-.72
I013	.81	.82	.88	.92	16.52	16.68	-.16
I014	.79	.80	.81	.84	16.24	16.36	-.12
I015	.77	.75	.74	.67	15.96	15.68	.28
I016	.79	.82	.81	.92	16.24	16.68	-.44
I017	.91	.94	1.34	1.55	18.36	19.20	-.84
I018	.75	.76	.67	.71	15.68	15.84	-.16
I019	.71	.72	.55	.58	15.20	15.32	-.12
I020	.87	.88	1.13	1.17	17.52	17.68	-.16
I021	.90	.93	1.28	1.48	18.12	18.92	-.80
I022	.95	.97	1.65	1.88	19.60	20.52	-.92
I023	.83	.89	.95	1.23	16.80	17.92	-1.12
I024	.84	.89	.99	1.23	16.96	17.92	-.96
I025	.94	.96	1.55	1.75	19.20	20.00	-.80
I026	.86	.89	1.08	1.23	17.32	17.92	-.60
I027	.79	.84	.81	.99	16.24	16.96	-.72
I028	.68	.75	.47	.67	14.88	15.68	-.80
I029	.67	.72	.44	.58	14.76	15.32	-.56
I030	.68	.73	.50	.61	15.00	15.44	-.44
I031	.84	.90	.99	1.28	16.96	18.12	-1.16
I032	.84	.90	.99	1.28	16.96	18.12	-1.16
I033	.80	.87	.84	1.13	16.36	17.52	-1.16
I034	.81	.88	.88	1.17	16.52	17.68	-1.16
I035	.80	.87	.84	1.13	16.36	17.52	-1.16
I036	.86	.90	1.08	1.28	17.32	18.12	-.80
I037	.80	.87	.84	1.13	16.36	17.52	-1.16
I038	.80	.82	.84	.92	16.36	16.68	-.32
I039	.85	.89	1.04	1.23	17.16	17.92	-.76
I040	.68	.68	.47	.47	14.88	14.88	.00
I041	.93	.95	1.48	1.65	18.92	19.60	-.68
I042	.97	.98	1.88	2.05	20.52	21.20	-.68
I043	.73	.71	.61	.55	15.44	15.20	.24
I044	.69	.71	.50	.55	15.00	15.20	-.20
I045	.85	.84	1.04	.99	17.16	16.96	.20
I046	.75	.71	.67	.55	15.68	15.20	.48
I047	.59	.56	.23	.15	13.92	13.60	.32
I048	.57	.54	.18	.10	13.72	13.40	.32
I049	.60	.58	.25	.20	14.00	13.80	.20
I050	.75	.72	.67	.58	15.88	15.32	.36
I051	.93	.93	1.48	1.48	18.92	18.92	.00
I052	.96	.98	1.75	2.05	20.00	21.20	-1.20
I053	.95	.97	1.65	1.88	19.60	20.52	-.92
I054	.92	.93	1.41	1.48	18.64	18.92	-.28
I055	.74	.71	.64	.55	15.56	15.20	.36
I056	.94	.96	1.55	1.75	19.20	20.00	-.80
I057	.93	.94	1.48	1.55	18.92	19.20	-.28
I058	.80	.81	.84	.88	16.36	16.52	-.16
I059	.74	.76	.64	.71	15.56	15.84	-.28
I060	.76	.80	.71	.84	15.84	16.36	-.52
I061	.77	.82	.74	.92	15.96	16.68	-.72
I062	.61	.73	.28	.61	14.12	15.44	-1.32
I063	.73	.75	.61	.67	15.44	15.68	-.24
I064	.91	.93	1.34	1.48	18.36	18.92	-.56
I065	.89	.94	1.23	1.55	17.92	19.20	-1.28
I066	.87	.91	1.13	1.34	17.52	18.36	-.84
I067	.89	.94	1.23	1.55	17.92	19.20	-1.28
I068	.70	.71	.52	.55	15.08	15.20	-.12
I069	.85	.89	1.04	1.23	17.16	17.92	-.76
I070	.85	.88	1.04	1.17	17.16	17.68	-.52

TOTAL RECORDS--GROUP 1 = 23851 : Males
TOTAL RECORDS--GROUP 2 = 22663 : Females

NOTE. Operational sequence numbers do not correspond to item numbers in actual test booklets due to the removal of field test items.

Table 5

Delta Values for Bias Analysis: Whites vs. Blacks

OPERATIONAL SEQUENCE NUMBER	P-VALUE G1	P-VALUE G2	Z-VALUE G1	Z-VALUE G2	DELTA G1	DELTA G2	DELTA (G1-G2)
I001	.97	.96	1.88	1.75	20.52	20.00	.52
I002	.96	.95	1.75	1.65	20.00	19.80	.40
I003	.99	.98	2.33	2.05	22.32	21.20	1.12
I004	.97	.94	1.88	1.55	20.52	19.20	1.32
I005	.96	.93	1.75	1.48	20.00	18.92	1.08
I006	.98	.98	2.05	2.05	21.20	21.20	.00
I007	.88	.83	1.17	.95	17.68	16.80	.88
I008	.95	.91	1.65	1.34	19.80	18.36	1.24
I009	.86	.75	1.08	.67	17.32	15.68	1.64
I010	.86	.78	1.08	.77	17.32	16.08	1.24
I011	.96	.90	1.75	1.28	20.00	18.12	1.88
I012	.88	.83	1.17	.95	17.68	16.80	.88
I013	.85	.77	1.04	.74	17.16	15.96	1.20
I014	.83	.74	.95	.84	16.80	15.56	1.24
I015	.78	.73	.77	.61	16.08	15.44	.64
I016	.84	.78	.99	.71	16.96	15.84	1.12
I017	.95	.90	1.65	1.28	19.60	18.12	1.48
I018	.81	.69	.88	.50	16.52	15.00	1.52
I019	.79	.60	.81	.25	16.24	14.00	2.24
I020	.91	.83	1.34	.95	18.36	16.80	1.56
I021	.93	.89	1.48	1.23	18.92	17.92	1.00
I022	.98	.95	1.75	1.65	20.00	19.60	.40
I023	.89	.82	1.23	.92	17.92	16.68	1.24
I024	.90	.82	1.28	.92	18.12	16.68	1.44
I025	.96	.93	1.75	1.48	20.00	18.92	1.08
I026	.91	.82	1.34	.92	18.36	16.68	1.68
I027	.87	.73	1.13	.61	17.52	15.44	2.08
I028	.74	.68	.64	.47	15.56	14.88	.68
I029	.73	.65	.61	.30	15.44	14.56	.88
I030	.75	.64	.67	.36	15.68	14.44	1.24
I031	.90	.81	1.28	.88	18.12	16.52	1.60
I032	.90	.81	1.28	.88	18.12	16.52	1.60
I033	.88	.76	1.17	.71	17.68	15.84	1.84
I034	.88	.79	1.17	.81	17.68	16.24	1.44
I035	.88	.77	1.17	.74	17.68	15.96	1.72
I036	.90	.84	1.28	.99	18.12	16.96	1.16
I037	.87	.78	1.13	.77	17.52	16.08	1.44
I038	.84	.74	.99	.64	16.96	15.56	1.40
I039	.90	.83	1.28	.95	18.12	16.80	1.32
I040	.74	.59	.64	.23	15.56	13.92	1.64
I041	.98	.91	1.75	1.34	20.00	18.36	1.64
I042	.99	.97	2.33	1.88	22.32	20.52	1.80
I043	.77	.63	.74	.33	15.96	14.32	1.64
I044	.75	.64	.67	.36	15.68	14.44	1.24
I045	.90	.78	1.28	.71	18.12	15.84	2.28
I046	.81	.60	.88	.25	16.52	14.00	2.52
I047	.63	.50	.33	.00	14.32	13.00	1.32
I048	.64	.40	.36	-.25	14.44	12.00	2.44
I049	.67	.47	.44	-.08	14.78	12.68	2.08
I050	.82	.60	.92	.25	16.88	14.00	2.88
I051	.95	.90	1.65	1.28	19.60	18.12	1.48
I052	.98	.98	2.05	1.75	21.20	20.00	1.20
I053	.97	.95	1.88	1.65	20.52	19.60	.92
I054	.95	.88	1.65	1.17	19.60	17.68	1.92
I055	.80	.60	.84	.25	16.36	14.00	2.36
I056	.97	.92	1.88	1.41	20.52	18.64	1.88
I057	.95	.91	1.65	1.34	19.60	18.36	1.24
I058	.87	.71	1.13	.55	17.52	15.20	2.32
I059	.82	.68	.92	.41	16.88	14.64	2.04
I060	.83	.68	.95	.50	16.80	15.00	1.80
I061	.84	.73	.99	.61	16.96	15.44	1.52
I062	.73	.57	.61	.18	15.44	13.72	1.72
I063	.80	.65	.84	.39	16.36	14.56	1.80
I064	.94	.90	1.55	1.28	19.20	18.12	1.08
I065	.93	.89	1.48	1.23	18.92	17.92	1.00
I066	.92	.84	1.41	.99	18.64	16.96	1.68
I067	.94	.89	1.55	1.23	19.20	17.92	1.28
I068	.78	.58	.77	.20	16.08	13.80	2.28
I069	.90	.82	1.28	.92	18.12	16.68	1.44
I070	.88	.82	1.23	.92	17.92	16.68	1.24

TOTAL RECORDS--GROUP 1 = 25501: Whites
TOTAL RECORDS--GROUP 2 = 10000: Blacks

NOTE. Operational sequence numbers do not correspond to item numbers in actual test booklets due to the removal of field test items.

error. Delta differences (G1 - G2) can be found in the far right column in Tables 4 and 5. All delta differences in both tables fall within the ±3.5 guidelines, indicating that no item on the 1987 MFMT-II shows statistical evidence of sex or race bias.

RESULTS

Pass-rates. The percentage of students who took the MFMT-II for the first time and passed improved slightly over 1986 (see Table 6 below). The pass-rates for the past six years for first-time takers as well as retested students are also presented.

Table 6

Pass-Rates for the MFMT-II: 1982-1987

	Percent Passing					
	School Year					
Grade	1982-83	1983-84	1984-85	1985-86	1986-87	1987-88
9	39.6*	60.8	61.0	64.9	65.5	68.3
10	NA	NA	33.6	41.8	40.2	44.4
11	NA	NA	NA	45.8	48.4	51.5
12	NA	NA	NA	NA	56.4	60.4

*Indicates a "no-fault" administration in which test performance does not affect graduation from high school, and for which no retesting occurs.

Note: Percentages include fall administration only.

NA: Not available.

Response Proportions. Table 7 contains the response proportions and p-values for each operational item on the MFMT-II. The response proportion indicates the proportion of students selecting each possible response for each item on the MFMT-II. Response proportions are used to evaluate the quality of the incorrect responses (or distractors), and can be useful in

Table 7

Percentages of Students Who Answered Each Item Correctly (P-Values) and Response Proportions

SEQUENCE NUMBER	RESPONSE PROPORTIONS A	B	C	D	OTHER	P-VALUE
I001	.02	.01	.01	.97	.00	.97
I002	.96	.01	.02	.01	.00	.96
I003	.99	.00	.00	.01	.00	.99
I004	.96	.01	.03	.01	.00	.96
I005	.02	.03	.95	.01	.00	.95
I006	.01	.98	.00	.01	.00	.98
I007	.11	.01	.87	.00	.00	.87
I008	.02	.93	.02	.02	.00	.93
I009	.14	.03	.01	.82	.00	.82
I010	.01	.83	.05	.11	.00	.83
I011	.04	.02	.93	.01	.00	.93
I012	.07	.05	.02	.86	.00	.86
I013	.03	.06	.82	.09	.00	.82
I014	.79	.04	.10	.07	.00	.79
I015	.03	.76	.11	.10	.00	.76
I016	.15	.03	.80	.02	.00	.80
I017	.92	.03	.01	.03	.00	.92
I018	.18	.76	.03	.03	.00	.76
I019	.71	.16	.10	.03	.00	.71
I020	.08	.87	.03	.02	.00	.87
I021	.91	.02	.02	.05	.00	.91
I022	.02	.96	.01	.01	.00	.96
I023	.07	.02	.05	.86	.00	.86
I024	.08	.05	.02	.87	.00	.87
I025	.01	.02	.95	.01	.00	.95
I026	.87	.08	.02	.02	.00	.87
I027	.08	.81	.11	.02	.00	.81
I028	.09	.03	.17	.71	.00	.71
I029	.69	.06	.12	.12	.00	.69
I030	.71	.11	.13	.04	.00	.71
I031	.87	.05	.06	.02	.00	.87
I032	.03	.04	.06	.87	.00	.87
I033	.05	.05	.06	.84	.00	.84
I034	.85	.03	.09	.03	.00	.85
I035	.06	.84	.07	.03	.00	.84
I036	.03	.08	.88	.01	.00	.88
I037	.05	.05	.06	.84	.00	.84
I038	.05	.81	.11	.03	.00	.81
I039	.87	.04	.04	.04	.00	.87
I040	.25	.68	.05	.02	.00	.68
I041	.00	.03	.94	.02	.00	.94
I042	.01	.98	.01	.01	.00	.98
I043	.10	.16	.02	.72	.00	.72
I044	.08	.70	.03	.19	.00	.70
I045	.05	.84	.06	.04	.00	.84
I046	.05	.73	.06	.16	.00	.73
I047	.20	.08	.58	.15	.00	.58
I048	.13	.56	.24	.07	.00	.56
I049	.24	.11	.05	.59	.00	.59
I050	.74	.23	.02	.01	.00	.74
I051	.05	.01	.93	.01	.00	.93
I052	.97	.02	.01	.01	.00	.97
I053	.01	.01	.03	.96	.00	.96
I054	.01	.04	.92	.02	.00	.92
I055	.22	.02	.03	.72	.00	.72
I056	.01	.02	.01	.95	.00	.95
I057	.04	.01	.93	.01	.00	.93
I058	.05	.81	.13	.01	.00	.81
I059	.75	.18	.02	.04	.00	.75
I060	.02	.13	.06	.78	.00	.78
I061	.03	.79	.09	.08	.00	.79
I062	.07	.16	.10	.67	.01	.67
I063	.74	.05	.05	.16	.01	.74
I064	.02	.03	.02	.92	.01	.92
I065	.02	.03	.01	.93	.01	.91
I066	.02	.89	.04	.04	.01	.89
I067	.03	.02	.02	.92	.01	.92
I068	.06	.14	.70	.09	.01	.70
I069	.02	.87	.04	.03	.01	.87
I070	.03	.03	.86	.08	.01	.86

TOTAL RECORDS = 46835

NOTE. Operational sequence numbers do not correspond to item numbers in actual test booklets due to the removal of field test items.

identifying confusing or miskeyed items. The p-value indicates the percentage of students who answered an item correctly. The p-values for the 1987 MFMT-II range from .56 (operational sequence number 48, actual item number 52) to .99 (operational sequence and actual item number 3). Experience has defined the acceptable range of p-values as .30 to .70 for norm-referenced tests. However, this convention applies less strictly to criterion-referenced tests that may include items outside of this range in order to ensure adequate coverage of relevant content areas. In particular, it is appropriate for basic skills tests like the MFMT-II to contain a number of "easier" items -- with p-values greater than .7 -- to ensure accurate measurement of low-achieving students and meaningful diagnostic domain scores.

Summary Results. The mean raw, domain, and scale scores and their standard deviations for the 1987 MFMT-II are presented in Table 8. The mean scale score of 352.8 represents a 1.3 point increase over 1986's mean scale score of 351.5. This gain is also reflected in the improvement in the 1987 pass-rate over 1986, as shown in Table 6. Finally, Table 9 presents the distribution of frequencies of each score point for the 1987 MFMT-II.

Table 8

Mean Raw, Domain, and Scale Scores and Standard Deviations (in parentheses): Grade 9 (N=46,635)

Domain	Maximum Raw Score	Raw Score	Domain Score	Scale Score
Number Concepts	10	8.2 (2.1)	80.3 (17.9)	353.2 (29.0)
Whole Number Operations	10	9.2 (1.2)	92.4 (9.2)	334.1 (22.1)
Mixed Number/ Fraction Operations	10	8.3 (2.4)	79.9 (22.6)	347.1 (30.8)
Decimal Operations	10	8.4 (1.9)	86.6 (14.3)	356.8 (29.6)
Measurement	10	7.4 (2.1)	79.1 (16.6)	353.2 (30.8)
Using Data	10	8.7 (1.7)	83.3 (16.2)	345.5 (26.4)
Problem Solving	10	8.3 (2.0)	81.3 (18.4)	349.1 (28.2)
All Domains	70	58.5 (10.6)	-- --	352.8 (29.3)

Table 9

Frequency Distribution of Total Scores

RAW SCORE	SCALE SCORE	PERCENT SCORE	FREQUENCY	FREQUENCY PERCENTAGE	CUMULATIVE FREQUENCY	CUMULATIVE PERCENTAGE
0	185	.0	3	.0	3	.0
1	205	1.4	0	.0	3	.0
2	219	2.9	0	.0	3	.0
3	228	4.3	0	.0	3	.0
4	235	5.7	3	.0	6	.0
5	240	7.1	2	.0	8	.0
6	245	8.6	0	.0	8	.0
7	249	10.0	3	.0	11	.0
8	252	11.4	5	.0	16	.0
9	256	12.9	1	.0	17	.0
10	258	14.3	4	.0	21	.0
11	261	15.7	5	.0	26	.1
12	264	17.1	9	.0	35	.1
13	266	18.6	8	.0	43	.1
14	268	20.0	13	.0	56	.1
15	271	21.4	24	.1	80	.2
16	273	22.9	24	.1	104	.2
17	275	24.3	30	.1	134	.3
18	276	25.7	43	.1	177	.4
19	278	27.1	43	.1	220	.5
20	280	28.6	57	.1	277	.6
21	282	30.0	82	.2	359	.8
22	284	31.4	62	.1	421	.9
23	285	32.9	70	.2	491	1.1
24	287	34.3	88	.2	579	1.2
25	288	35.7	105	.2	684	1.5
26	290	37.1	105	.2	789	1.7
27	292	38.6	101	.2	890	1.9
28	293	40.0	127	.3	1017	2.2
29	295	41.4	132	.3	1149	2.5
30	296	42.9	131	.3	1280	2.7
31	298	44.3	173	.4	1453	3.1
32	299	45.7	176	.4	1629	3.5
33	301	47.1	187	.4	1816	3.9
34	302	48.6	207	.4	2023	4.3
35	304	50.0	210	.5	2233	4.8
36	305	51.4	211	.5	2444	5.2
37	306	52.9	229	.5	2673	5.7
38	308	54.3	255	.5	2928	6.3
39	309	55.7	263	.6	3191	6.8
40	311	57.1	315	.7	3506	7.5
41	312	58.6	333	.7	3839	8.2
42	314	60.0	370	.8	4209	9.0
43	315	61.4	373	.8	4582	9.8
44	317	62.9	426	.9	5008	10.7
45	318	64.3	444	1.0	5452	11.7
46	320	65.7	491	1.1	5943	12.7
47	321	67.1	575	1.2	6518	14.0
48	323	68.6	603	1.3	7121	15.3
49	325	70.0	704	1.5	7825	16.8
50	326	71.4	743	1.6	8568	18.4
51	328	72.9	836	1.8	9404	20.2
52	330	74.3	911	2.0	10315	22.1
53	332	75.7	991	2.1	11306	24.2
54	334	77.1	1009	2.2	12315	26.4
55	336	78.6	1125	2.4	13440	28.8
56	338	80.0	1297	2.8	14737	31.6
57	340	81.4	1374	2.9	16111	34.5
58	342	82.9	1547	3.3	17658	37.9
59	344	84.3	1568	3.4	19226	41.2
60	347	85.7	1741	3.7	20967	45.0
61	350	87.1	1944	4.2	22911	49.1
62	353	88.6	2095	4.5	25006	53.6
63	356	90.0	2249	4.8	27255	58.4
64	360	91.4	2494	5.3	29749	63.8
65	364	92.9	2646	5.7	32395	69.5
66	369	94.3	3039	6.5	35434	76.0
67	376	95.7	3086	6.6	38520	82.6
68	385	97.1	3285	7.0	41805	89.6
69	399	98.6	2945	6.3	44750	96.0
70	418	100.0	1885	4.0	46635	100.0

NUMBER OF RECORDS = 46635

References

Angoff, W.H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), Handbook of methods for detecting item bias. Baltimore: The Johns Hopkins University Press.

Chafin, A.E. (1983). Setting a standard for standard setting. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. NY: CBS College Publishing.

Gronlund, N. E. (1982). Constructing achievement tests (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Maryland State Department of Education (1980). Declared competencies index. Baltimore, MD: Author.

Maryland State Department of Education. (1981). Project basic instructional guide, Volume V: Functional mathematics. Baltimore, MD.

Subkoviak, M. (1984). Estimating the Reliability of Mastery-Nonmastery Classifications. In R. A. Berk (Ed.), A guide to criterion-referenced test construction. Baltimore: The Johns Hopkins University Press.

LWB16000